DOCUMENT RESUME

ED 178 569 TM 009 714

AUTHOR Veitch, William R.
TITLE An Application of the Rasch Model.
PUB DATE Mar 79
NOTE 10p.; Paper presented at the annual meeting of the Michigan School Testing Conference (Ann Arbor, MI, March, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cutting Scores; Educational Assessment; Intermediate Grades; *Item Analysis; Mastery Tests; Mathematical Models; Reading Tests; *Scoring Formulas; *Test Interpretation
IDENTIFIERS Michigan Educational Assessment Program; *Rasch Model

ABSTRACT

The one parameter latent trait theory of Georg Rasch has two assumptions: that student abilities can be measured on an equal interval scale, and that the success of a student with a given item is a function of student achievement and item difficulty. The grade four Michigan Educational Assessment Program reading test was designed to measure nineteen objectives with 95 items (five items for each objective). Mastery for each objective was defined as four correct out of five. The data analysis for this study consisted of separately analyzing the total score on the 95 items and the mastery scores on the 19 objectives, and placing both scores on the same Rasch model score scale. A table is presented showing the raw test score ranges for those students with various mastery scores. The author concludes that Rasch scaled scores would have provided more information than the mastery approach. (CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
William R. Veitch
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED178569

TM009 714

ABSTRACT

An Application of the Rasch Model

William R. Veitch

Item characteristic curve theory or latent trait analysis (popularly called Rasch Modeling) has two basic assumptions in the one-parameter model. First is that academic ability for a given trait can be measured on an equal interval scale. The second assumption is that the degree of success for a given student is a function of student ability and test item difficulty. The Rasch Model's mathematics produce item diffculty estimates that are stable across groups with widely varying abilities.

This application describes the efficacy of the use of the Rasch equal interval scale score for total test as opposed to the use of percent of objectives mastered that is common in today's objective referenced testing environment.

An Application of the Rasch Model[1]

William R. Veitch
Oakland Schools

An area of psychometric research variously known as "item characteristic curve theory", or "latent trait analysis", and popularly referred to as "Rasch modeling" has been making inroads into the testing literature over the past ten years. Latent trait analysis comes in several forms: the one, two, and three parameter models. The most complex model, Birnbaum's (1968) three parameter, attempts to take into account for any given set of items (1) item difficulty, (2) item discrimination, and (3) a correction for guessing. The two parameter model eliminates the guessing parameter, while the one parameter (Rasch) model operates only with estimates of the item difficulties.

INTRODUCTION

The Rasch model is named after Danish mathematician Georg Rasch (1960,1966). It was Rasch who provided the psychometric community with the straightforward and tractable theory and mathematics for the independent estimation of student achievement and item difficulty. Latent trait theory in the United States has found a major proponent in Benjamin D. Wright (see Wright, 1977)

[1]Paper presented at the Michigan School Testing Conference, Ann Arbor, Michigan, March, 1979.

TM009 714

of the University of Chicago. The Rasch approach involves two basic assumptions. The first assumption is that an academic ability for a given trait can be measured on an equal interval scale. The concept of this measurement of achievement goes beyond the idea of a measure on a particular test. It may be viewed in the same light as the "true score" from traditional statistics. That is, the assumption uses the equal interval scale score (1) as a measure of a student's "true" achievement in the trait area (with any set of items from the trait area), and (2) to estimate item parameters (difficulty, etc.) independently from the level of performance of the sample.

The second basic assumption is that the degree of success a student will have with a given item is a function of student achievement and item difficulty. In other words: able students do well on all items except the hardest, and less able students do well only on easy items. These two assumptions and a considerable amount of powerful mathematics are combined to produce (from simple p-values and raw scores) an independent equal interval scale which measures item difficulty and student achievement. The traditional p-value will vary greatly from group to group, but Rasch's mathematics produce item difficulties that are stable even between groups of widely varying achievement. The appeal of this characteristic is that it virtually eliminates the need

for random samples when field testing. It also presents the psychometrician with the possibility of measuring student progress against a curriculum, rather than against national norms.

APPLICATION

The data for the analysis described in this paper are from the grade four Michigan Educational Assessment Program reading test as administered in the Fall of 1977 in a mid-sized Southeast Michigan school district. The test was comprised of nineteen objectives with five items each for a total of 95 items. Although the MEAP test contains 95 items, it is more generally viewed as a 19 item test. Mastery of an objective has been defined as 4 correct out of 5, and most student achievement is thought of in terms of the percent of objectives mastered.

The standard procedure for Rasch analysis of a group of items is to calibrate (generate item difficulty estimates) all items on all subjects (n = 366 in this case). The second step is to plot individual item responses so as to identify persons who do not "fit" the test (see Mead, 1976). The final step is to re-calibrate the total test on all persons who "fit". At this point it would be appropriate to delete those items that do not fit the test so as to yield the "best" test possible.

In Table 1, the reader will find the results of jointly calibrating grade 4 MEAP reading as a 95 and as a 19 item test. The first two columns report total number

of items correct (i.e., raw score) and the associated standardized achievement score (mean = 200, stdv = 10). The next two columns report numbers of objectives mastered and their standard scores. Each objective score is placed at the minimum point on the raw score scale at which that number of objectives could have been mastered (e.g., 18 obj. X 4 items = 72 raw score). Column five reports the actual range of raw scores that persons received who passed that particular number of objectives.

A careful inspection of Table 1 will reveal that there is considerable discrepancy between total raw score and number of objectives mastered as measures of student achievement. The Rasch calibration has converted each of the scores to the common scale that allows this comparison. For example, for a person to pass one objective would theoretically require a total raw score of only 4, all items passed on the one objective mastered. On the other hand, another individual might pass 3 of 5 items on all 19 objectives for a raw score of 57 and master no objectives. Table 1 shows that a raw score of 4 is a standard score of 167, while a raw score of 57 converts to a standard score of 204, a discrepancy of nearly four standard deviations. The reader will note that the actual range of raw scores associated with passing one objective was from 7 to 42 (173 to 197), and the maximum raw score for no objectives mastered was 32 (193). The bottom line is the question of which of the

available measures (raw score vs. number of objectives mastered) yields to the educator the more accurate and usable information on student achievement levels.

DISCUSSION

Consider the fact that a student in the data sample correctly answered 32 items on grade 4 MEAP, and on the basis of percent mastery was adjudged to have zero reading abilities. A score of 32 (and it could have been as high as 57) is well above chance level on this test, and indicates at least some glimmering of reading comprehension. The standard score associated with 32 (193) indicates an individual at that level to be only one and one-half standard deviations below the mean (208) of this sample group. Much worse, a raw score of 57 (204) would actually be only four points below the mean. Given the current reporting scheme, it is possible for students to achieve such Rasch standard scores with zero mastery in the objective referenced sense.

The crux of the matter is the attempt to give the educator more information. Nineteen onjectives have been identified for the grade 4 reading test and are therefore reported for each student. That seems like more information than one total raw score. However, what at first seems like an attempt to provide increased information, actually winds up as a loss. Remember that the "score" on state assessment is percent mastery of objectives. By dichotomizing 19 objectives into pass/fail

instead of 95 items into right/wrong, we lose information on student achievement at the rate of an 80% reduction. Again, the comparison of standard scores from Table 1 confirms this. The arbitrary criterion of 4 correct out of 5 for objective mastery, and the reporting of student "scores" in terms of percent mastery is costing the educator valuable psychometric information. It is costing the educator the more accurate estimate of student achievement provided by the Rasch model's standard scale score.

REFERENCES CITED

BIRNBAUM, A. Some latent trait models and their use in inferring an examinee's achievement. In F.M. Lord & M.R. Novick, Statistical Theories of Mental Test Scores. Reading, Mass.: Addison-Wesley, 1968.

MEAD, R. Assessing the fit of data to the Rasch model through analysis of residuals. Doctoral dissertation, University of Chicago, 1976.

RASCH, G. Probabilistic models for some intelligence and attainment tests. Copenhagen, Denmark: Paedogogiske Institut, 1960.

RASCH, G. An item analysis which takes individual differences into account. British Journal of Mathematical and Statistical Psychology, 1966, 19(1), 49-57.

WRIGHT, B. Solving measurement problems with the Rasch Model. Journal of Educational Measurement, 1977, 14(2) 97-116.

TABLE 1
JOINT CALIBRATION OF MEAP AS 55 & 19 ITEM TEST

Raw Score	Standard Score	Obj. Mastered	Standard Score	Raw Score Range
94	247	.	.	.
93	240	.	.	.
92	236	.	.	.
91	233	.	.	.
90	231	.	.	.
89	229	.	.	.
88	227	.	.	.
87	226	.	.	.
86	224	.	.	.
85	223	.	.	.
84	222	.	.	.
83	221	.	.	.
82	220	.	.	.
81	219	.	.	.
80	218	.	.	.
79	217	.	.	.
78	217	.	.	.
77	216	.	.	.
76	215	19	295	(82-95)
75	214			
74	214			
73	213			
72	212	18	242	(82-93)
71	212			
70	211			
69	211			
68	210	17	235	(75-92)
67	210			
66	209			
65	208			
64	208	16	230	(75-90)
63	207			
62	207			
61	206			
60	206	15	226	(71-84)
59	205			
58	205			
57	204			
56	204	14	223	(71-83)
55	203			
54	203			
53	203			
52	202	13	220	(67-83)
51	202			
50	201			
49	201			
48	200	12	217	(65-77)
47	200			
46	199			
45	199			
44	198	11	214	(63-73)
43	198			
42	197			
41	197			
40	196	10	212	(57-73)
39	196			
38	196			
37	195			
36	195	9	209	(54-72)
35	194			
34	194			
33	193			
32	193	8	206	(50-68)
31	192			
30	192			
29	191			
28	190	7	204	(45-65)
27	190			
26	189			
25	189			
24	188	6	201	(43-62)
23	188			
22	187			
21	186			
20	186	5	198	(33-62)
19	185			
18	184			
17	183			
16	183	4	194	(24-57)
15	182			
14	181			
13	180			
12	179	3	190	(14-52)
11	178			
10	177			
9	176			
8	174	2	184	(13-50)
7	173			
6	171			
5	169			
4	167	1	176	(7-42)
3	164			
2	160			
1	153			
0	113	0	115	(0-32)

Discrepancy Range for 1 Obj.

Max = (18 x 3) + 4 = 58 = 205

Min = (1 x 4) = 4 = 167

Discrepancy Range for 18 Obj.

Max = (18 X 5) + 3 = 93 = 240

Min = (18 X 4) = 72 = 212